DOCUMENT RESUME

ED 223 640 TM 820 669

AUTHOR Jones, Douglas H.

TITLE Redescending M-Type Estimators of Latent Ability. Program Statistics Research, Technical Report No. 82-30.

INSTITUTION Educational Testing Service, Princeton, N.J.

REPORT NO ETS-RR-82-15

PUB DATE Mar 82

NOTE 22p.

PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Estimation (Mathematics); *Guessing (Tests); *Latent Trait Theory; *Mathematical Models; Research Methodology; Statistical Analysis; Statistics; Test Theory

ABSTRACT

New ability estimators have been proposed by Wainer and Wright (1980) and Mislevy and Bock (1981) that are resistant against guessing and careless behaviors exhibited by some examinees. This paper presents another class of ability estimators that are resistant to departures from the underlying assumptions concerning guessing and carelessness. In addition to computing the asymptotic relative efficiency of such estimators, this paper evaluates estimators by comparing their influence curves (Huber, 1981). (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223640

RR-82-15

Redescending M-Type Estimators of Latent Ability

Douglas H. Jones

PROGRAM STATISTICS RESEARCH

TECHNICAL REPORT NO. 82-30

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. H. Jones

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY 08541

TM 820 669

Redescending M-Type Estimators of Latent Ability

Douglas H. Jones

Program Statistics Research
Technical Report No. 82-30

ETS Research Report 82-15

March 1982

This research was supported by the Program Statistics Research Project of the Educational Testing Service Research Statistics Group.

Copyright © 1982 by Educational Testing Service. All rights reserved.

The Program Statistics Research Technical Report Series is designed to make the working papers of the Research Statistics Group at Educational Testing Service generally available. The series consists of reports by the statisticians in the Research Statistics Group as well as their external and visiting statistical consultants.

Reproduction of any portion of a Program Statistics Research Technical Report requires the written consent of the author(s).

ERIC

Abstract

New ability estimators have been proposed by Wainer and Wright (1980) and Mislevy and Bock (1981), that are resistant against guessing and careless behaviors exhibited by some examinees. This paper presents another class of ability estimators that are resistant to departures from the underlying assumptions concerning guessing and carelessness. In addition to computing the asymptotic relative efficiency of such estimators, this paper evaluates estimators by comparing their influence curves (Huber, 1981).

ERIC

Table of Contents

ERIC

1. Introduction

New ability estimators have been proposed by Wainer and Wright (1980) and Mislevy and Bock (1981), that are resistant against guessing and careless behavior exhibited by some examinees. This paper presents another class of ability estimators that are resistant to departures from the underlying assumptions concerning guessing and carelessness. In addition to computing the asymptotic relative efficiency of such estimators, this paper evaluates the estimators by comparing their influence curves (Huber, 1981).

It is of some importance to note two difficulties that have had to be overcome in the derivation of the asymptotic behavior of the estimators. The first is that the desired results do not follow directly from those for maximum likelihood estimators since the new class of estimators include some estimators that are not maximum likelihood. In fact, the asymptotic behavior has been derived from first principles. The second difficulty is that item responses are not identically distributed random variables when the items differ in difficulty, discrimination, or guessing characteristics. This has been overcome by assuming the items to be randomly sampled from a parent population.

2. Definition and Motivation of a New Class of Estimators

Let $x_1, \ldots, x_n$ be independent dichotomous item responses such that for a candidate with ability T, a real value parameter, $\Pr(X_i = x_i | T) = [P_i(T)]^{x_i}[1-P_i(T)]^{1-x_i}$, $x_i = 1$ or 0 where the $P_i(\cdot)$'s, called item response curves, are possibly different mappings from the real line to the unit interval [0, 1]. For convenience the subscript i will be deleted.

The proposed estimator of T is defined as the solution to the equation

$$\Sigma a[x - P(t)][P(T)Q(T)]^{h-1}=0 \qquad (2.1)$$

where the sum is over all items, Q=1-P, the a's are given but possibly different constants, and h is a real number greater than or equal to 1. In the foregoing we refer to these estimators as h-estimators.

The value of h is chosen according to how much robustness is desired; the greater the value of h, the more robust the estimator is. Guidelines for the choice of h depend on the value of the estimator's asymptotic variance that can be tolerated in order to reduce the influence of individual responses on the estimator. More discussion on this topic will follow in the next several sections.

Under certain circumstances, h-estimators correspond to maximum likelihood estimators (mle's). If, in addition to the above assumptions, $dP(t)/dt = \dot{P}(t)$ exists for each item, the mle is the solution to

$$\Sigma\dot{P}(T)[x-P(T)][P(T)Q(T)]^{-1}=0 .$$

Furthermore, if all P satisfy the two-parameter logistic model:

$$\ln[P(t)/Q(t)] = a(t-b),$$

then

$$\dot{P}(t) = aP(t)Q(t).$$

So for h = 1, the h-estimator is the mle for the two-parameter logistic model.

Both h-estimators and mle's are special cases of more general kinds of estimators: those that are solutions to

$$w(T)[x-P(T)] = 0,$$

for some function w.

For various kinds of w-functions, we have:

1) An h-estimator is the special case

$w(t) = a[P(t)Q(t)]^{h-1}$;

2) An mle arises when

$w(t) = d\ln[P(t)/Q(t)]/dt$.

We will denote the weight functions of the h-estimators by

$$w(t;a,h) = a[P(t)Q(t)]^{h-1} \tag{2.2}$$

A priori, h-estimators with reasonable weight functions should possess good robustness properties. First, they should not be overly influenced by any one item response and second, they should be stable when the true model for response departs from the assumed mode for response. Reasons for these assertions are discussed as follows.

<u>h-estimators should resist the influence of single outlying responses</u>

Suppose a new item is administered and that h is greater than one. If T_n, based on n responses, is already such that $P_{n+1}(T_n)$ is near 0, but $x_{n+1}=1$, then $[x_{n+1}-P_{n+1}(T)]w(T;a_{n+1},h)$ is a relatively small contribution to the sum in (2.1) defining the new estimator T_{n+1}. So the new observation will not dramatically change the old estimate of T. A similar finding holds for $P_{n+1}(T_n)$ near 1, but $x_{n+1} = 0$, since the sum in (2.1) is symmetric in the value P_{n+1}.

<u>Example 1</u> Suppose the model for item response is the logistic model with b_i =-0.8 to 1.0 by steps of 0.2 and a_i=1. Table 1 displays the various values of the h-estimator for two item response sequences: one without an outlier and one with an outlier. The h-estimators are less

affected by the outlier than the mle, (h=1). Further, the effect is smaller, the larger the value of h.

h-estimators should be insensitive to departures from the model

Suppose the true model is P* unequal to P for each item. One could retain the old weight function so that the estimator remains resistant to outliers, but solve

$$\Sigma[x-P^*(T)]\, w(T;a,h) = 0 \tag{2.3}$$

for a reasonable estimator of ability. The equation (2.1) is equal to the above equation plus the term

$$[P^*(T)-P(T)]\, w(T;a,h)$$

added to its right hand side. If this term is small, then solutions to (2.1) are close to solutions to (2.3). Since this term gets smaller as h gets larger, we expect h-estimators to be robust to departures from the model.

In fact, the property described is precisely continuity of the estimator when viewed as a function of the P_i's. We show in the foregoing that h-estimators are continuous functions under the proper mathematical setting.

3. The Influence Function and Other Heuristics

The influence function is a useful tool in robust statistics. Not only does it allow the evaluation of the influence of outliers on the estimator under investigation, it also allows a heuristic derivation of the limiting law of its sample distribution function (Hampel, 1968, 1974: Huber, 1981). Of course, the result must be checked with a rigorous mathematical proof.

The Influence Function of the h-estimator

For general statistical estimators, the derivation of the influence function is facilitated by viewing the estimator as a function of the probability distribution function F of the observations. Then the influence function is derived from the Gateaux derivative response (Huber, 1981) of the estimator function, T(F), at a distribution function F in the direction G:

$$\dot{T}(G;F) = \lim_{s\to o} \frac{T((1-s)F+sG)-T(F)}{s} .$$

The last expression is the ordinary derivative of $T((1-s)F+sG)$ with respect to s. (Further discussion and references will be found in Huber (1981) pages 13 and 37.) The influence function at z_o is the Gateaux derivative in the direction, $G(z) = d(z, z_o)$, which is the point-mass at z_o.

The difficulty in determining the influence function of item ability estimators lies in representing them as functions of the probability distribution functions of the observations. The defining equation (2.1) relates each value of the estimator to every set of values of item responses, item response curves, and constants: $z_i = (x_i, P_i, a_i)$, i=1, ..., n. Denote the point-mass at z_i as above, and define the empirical distribution function as $F_n(z)=\Sigma d(z, z_i)/n$. Clearly, the estimator defined in (2.1) depends on F_n since (2.1) is equivalent to

$$\int [x-P(T)]\, w(T;a,h)\,dF_n=0. \qquad (3.1)$$

Denote this dependence by the functional notation $T=T(F_n)$. The function $T(\cdot)$ will be extended to any probability mass function F by replacing F_n by F in (3.1).

$$\int [x-P(T(F))]\, w(T(F);a,h)\,dF=0. \tag{3.2}$$

Since the functional notation is defined implicitly, substitution of $(1-s)\,F+sG$ for F in (3.2) and chain-rule differentiation with respect to s yields an equation involving the Gateaux derivative:

$$\dot{T}\,\frac{d}{dt}\int [x-P(t)]\, w(t;a,h)\,dF\Big|_{t=T} \tag{3.3}$$

$$+\int [x-P(T)]\, w(T;a,h)\,(dG-dF)=0.$$

where $T=T(F)$. Because (3.2) is satisfied, letting $G=d(z,z_i)$ in (3.3) yields the following influence function:

$$IC\,(z_i;F,T) = \frac{[x_i-P_i(T)]\,w_i(T,a_i,h)}{\frac{d}{dt}\int [x-P(t)]\, w(t;a,h)\;dF\Big|_{t=T}} \tag{3.4}$$

(The notation IC refers to Influence Curve). Where the subscript for w is added to emphasize its dependence on the i-th response curve.

Comparison between the influence curve of estimators for different values of T are useful. The usual notion of an influence function is a curve in x; however, since x takes only two values this notion is not useful for item response theory. On the other hand, these influence functions are curves in T;

however, graphs of the influence function are not as easily manufactured as they are for the problems typically investigated by statisticians. The difficulty lies in the dependence of the denominator on the value of T.

A few examples of the influence curves are plotted in Figure 1 as functions of F(T). The general form of the curve for h strictly greater than one, excluding the mle, approach zero asymptotically as T approaches infinity. This indicates that outlying responses have less influence for very large or very small abilities. The behavior of the influence function for h equal to one, the mle, reveals that the largest influences are obtained for T approaching negative or positive infinity.

Compared to the mle, the influence function of the h-estimator is redescending; that is, for either response, the influence starts at zero, rises and returns to zero. Hence, the property of influence functions of h-estimators are analogous to those of redescending M-estimators for the location problem in standard statistics.

Conjectured Asymptotic Normality of h-estimators

If F_n has the limiting value F, then the one-term Taylor expansion entails

$T(F_n) = T(F) + \Sigma IC(Z_i; F,T) + R(F_n,F)$, where $R(F_n,F)$ is the reaminder term depending on F_n and F. We have denoted Z_i to be a random triplet with distribution F. We would expect, for heuristic reasons, that $n^{\frac{1}{2}}R(F_n,F)$ converges to zero. Consequently, $n^{\frac{1}{2}}[T(F_n) - T(F)]$ would have the same limiting value of

$n^{\frac{1}{2}}\Sigma IC(Z_i;F,T)$. We state the consequences of these results as a theorem.

Theorem If the response and items are sampled so that Z_i are independent replicates from F and $n^{\frac{1}{2}}R(F_n,F)$ converges to zero then $n^{\frac{1}{2}}[T(F_n) - T(F)]$ has a limiting normal distribution with mean 0 and variance

$$A(F,T) = \frac{\int a^2[x-P(T)]^2[P(T)Q(T)]^{2(h-1)}dF}{[\frac{d}{dt}\int a[x-P(t)]P(t)Q(t)^{h-1}dF\big|_{t=T}]^2} . \qquad (3.5)$$

In the following, this theorem will be proven with sufficient conditions weaker than $n^{\frac{1}{2}}R(F_n,F)$ converging to zero.

4. Asymptotic Properties

We will describe a probability structure that conveniently yields the asymptotic behavior of the ability estimators in item response theory. Let Z be the set of all triples $z=(x,P,a)$ as x=0 or 1, P ranges through a finite set P of non-decreasing maps from the entire real line to the closed unit interval, and a ranges through a positive finite set A. The set Z, imbued with the discrete topology on its power set, is the sample space. Distributions over Z are defined as follows. If $\{f(z), z\varepsilon Z\}$ is a set of positive weights, summing to one, for a subset B, the associated distribution function F on Z is expressed as $F(B)=\Sigma_{z\varepsilon B}f(B)$ Let $z_1, \ldots, z_n$ be a random sample from F, then the empirical distribution function is $F_n(B)=\Sigma_{i=1,n}d(z_i,B)/n$, where $d(z,B)=1$ if $z\varepsilon B$ and 0 otherwise. F_n enjoys several properties following from the Strong

Law of Large Numbers and the Central Limit Theorem: a) for any real valued function g on Z, $\int g(z)dF_n(z)$ converges to $\int g(z)dF(z)$ wpl, and b) if $\int g^2(z)dF$ is finite, $n^{\frac{1}{2}}[\int g(z)dF_n(z) - \int g(z)dF(z)]$ is asymptotically normal with mean zero and variance $\int [g(z)-\bar{g}]^2 dF(z)$ where $\bar{g} = \int g(z)dF(z)$. The latter integration is the expectation of g with respect to F denoted by $E_F g$. In the foregoing we consider $g(z)=g(z;t)=a[x-P(t)][P(t)[1-P(t)]]^{h-1}$.

Define $m_F(t) = E_F g(Z;t)$. We assume throughout that $Z_1,\ldots,Z_n$ is a random sample from F with empirical distribution function F_n.

Theorem 1 (Consistency) Let t_o be an isolated root of $m_F(t) = 0$. Suppose that P(t) is continuous in t for each P in $\mathcal{P}$. If T_n is a solution sequence to the empirical equation $m_{F_n}(T_n) = 0$, then T_n converges to t_o wpl.

Proof of Theorem 1 Since P(t) is continuous, $m_F(t)$ is continuous in t. Therefore, for each ε sufficiently small $m_F(t_o - \varepsilon)$ and $m_F(t_o + \varepsilon)$ are opposite in sign. Without loss of generality assume $m_F(t_o - \varepsilon)<0< m_F(t_o+ \varepsilon)$.

Since g(Z;t) is bounded in z for each t, the Strong Law of Large Numbers implies $m_{F_n}(t) \to m_F(t)$ wpl. Hence

$$\lim n \to\infty P\{m_{F_n}(t_o - \varepsilon) < 0 < m_{F_n}(t_o+ \varepsilon), \text{ for all } m>n\} = 1.$$

But this implies

$$\lim n \to\infty P\{t_o - \varepsilon < T_m < t_o+\varepsilon, \text{ for all } m>n\} = 1.$$

The proof is complete.

Remark 1 Theorem 1 is valid even for a model of item responses that is different from the assumed model. The simplest way to see this is to allow $F=F^*$ where F^* induces a random variable P^* in $\mathcal{P}$ such that $E_{F^*}(X|P^*)=P^*$.

Remark 2 Even if t_o does not correspond to T, the true ability, the solution sequence T_n will converge t_o, and t_o can not correspond to T unless $m_{F^*}(T)=0$.

The final remark brings us to our next theorem about the magnitude of the difference between the true ability T and the limit of a sequence of estimators under an alternative model. Let F^* be a distribution over Z induced by a mapping from $\mathcal{P}$ into $\mathcal{P}$ denoted by P^*. The item response curve alternative to P is $P^*(P)$ and its value at t is denoted by $P^*(P)(t)$. Define $m^*_F(t) = E_F a[X-P^*(P)(t)][P(t)Q(t)]^{h-1}$. Define the true ability to be a solution to $m^*_F(t)=0$.

Theorem 2 (Asymptotic Bias) Let t_o be a solution to $m_F(t) = 0$ and let T be a solution to $m^*_F(t) = 0$. Then

$$T-t_o \doteq m_F(T)/[dm_F(t)/dt\big|_{t=t_1}]$$

where $|t_1-t_o| < |T-t_o|$.

Proof The Mean Value Theorem implies $m_F(T)-m_F(t_o) = (T-t_o)\, dm_F(t)/dt\big|_{t=t_1}$ where $|t_1-t_o| < |T-t_o|$. The theorem follows from the definition of t_o.

The next corollary insures that any solution sequence converges with probability one. This is a strong result that indicates how strong the standard assumptions of item response theory are.

Corollary 1 Suppose $P(t)$ is continuous in t for each P in $\mathcal{P}$. Assume that for each P in $\mathcal{P}$, P is a cdf in t. If T_n is a solution to the empirical equation $m_{F_n}(T_n) = 0$, then T_n converges wp1.

Proof We must show that $m_{F^*}(t) = 0$ always has a solution for any F^* that is the limit of F_n. Suppose that $m_{F^*}(t) \neq 0$ for any t. Then F^* gives positive probability to a set of P^* and P such that $P^*(T)-P(t) \neq 0$ for all t. Since $0 < P^*(T) < 1$, the last condition leads to a contradiction to the assumed continuity of each P.

We now turn to the distributional properties of a solution sequence.

Theorem 3 (Asymptotic Normality) Let t_o be an isolated root of $m_F(t) = 0$. Suppose that $dP(t)/dt$ is continuous in t uniformly in $\mathcal{P}$. Let T_n be a solution sequence of $m_{F_n}(t) = 0$ satisfying $T_n \to t_o$. Then T_n is asymptotically normal with variance $A(F, t_o)$ given by

$$A(F,t_o) = Eg^2(Z;t_o)/[Eg'(Z;t_o)]^2 .$$

Proof of Theorem 3 For notational convenience define $u_i(t) = g(Z_i;t) = a_i[X_i-P_i(t)][P_i(t)(1-P_i(t)]^{h-1}$. Since $dP_i(t)/dt$ is continuous, we may apply the Mean Value Theorem to obtain the expansion

$$u_i(T_n)-u_i(t_o) = (T_n-t_o) \; du_i(t)/dt\Big|_{t=t_n}$$

where $|t_n-t_o|<|T_n-t_o|$. Since $m_{F_n}(T_n)=0$, we have

$$n^{\frac{1}{2}} (T_n-t_o) = \frac{n^{\frac{1}{2}} \Sigma u_i(t_o)}{n^{-1} \Sigma \frac{d}{dt} u_i(t) \Big|_{t=t_n}}$$

The Central Limit Theorem implies that the numerator is asymptotically normal with mean 0 and variance $E_F g^2(Z,t_o)$. The Strong Law of Large Numbers, the hypothesis that $dP(t)/dt$ is uniformly continuous, and $T_n \to t_o$ imply that the denominator converges to $E_F g'(Z;t_o)$ wp1. The proof follows from Slutsky's Lemma.

We will use these theorems to compare estimators in the next section.

5. Efficiency Comparisons

Recall that when h=1, the h-estimator corresponds to a certain mle associated with a two-parameter logistic model having discrimination parameters equal to the corresponding values of a_i that appear in (2.1). Consequently, it is easy to determine the asymptotic efficiency of the h-estimator relative to this mle.

The first comparison is made under the associated two-parameter logistic model and the second comparisons are made under a neighboring two-parameter logistic model. The first model corresponds to the one appearing in Example 1. The sampling scheme consists of choosing from the designated items with equal probability.

Table 2 displays the asymptotic relative efficiencies when the true F* corresponds to the assumed model. For the computed values of h, the h-estimators lose no more than 10 percent efficiency, or one in ten items is wasted.

Table 3 displays the asymptotic relative efficiency when the true F* generates a $P^*(t)=P(t-0.1)$ for each sampled item. This is equiva-

lent to uniformly shifting the difficulty parameters to the left by 0.1, or 6 percent of the total range of the difficulty parameters. In computing the efficiency we have used the approximation to the asymptotic bias given by Theorem 2 in order to compare the mean squared errors. The h-estimators outperform the mle for each computed value of h.

6. References

Hampel, F.R. Contributions to the theory of robust estimation. Ph.D. Thesis, University of California, Berkeley.

_______ The influence curve and its role in robust estimation. Journal of the American Statistical Association, 62, pp. 1179-1186, 1974.

Huber, P.J. Robust Statistics New York: John Wiley, 1981.

Lord, F. and Novick, M. Statistical Theory of Mental Test Scores Reading, Massachusetts: Addison-Wesley, 1968.

Mislevy, R.J. and Bock, R.D. Biweight estimates of latent ability. Manuscript, 1981.

Serfling, R.J. Approximation Theorems of Mathematical Statistics New York: John Wiley, 1980.

Slutsky, E. Über stochastiche Asymptoter und Grenqwerte, Math. Annalen, 5, p. 93, 1925.

Wainer, H. and Wright, B.D. Robust estimation of ability in the Rasch Model. Psychometrika, 45, pp. 373-390, 1980.

7. Tables

Table 1: Values of the h-estimator at chosen response sequences

Responses	h				
X	1(MLE)	1.5	2	3	4
1111100000	0.11	0.11+	0.14	0.13	0.12
1111100001	0.58	0.41	0.22	0.20	0.19

Table 2: Efficiency comparison of h-estimators to the MLE under the assumed model

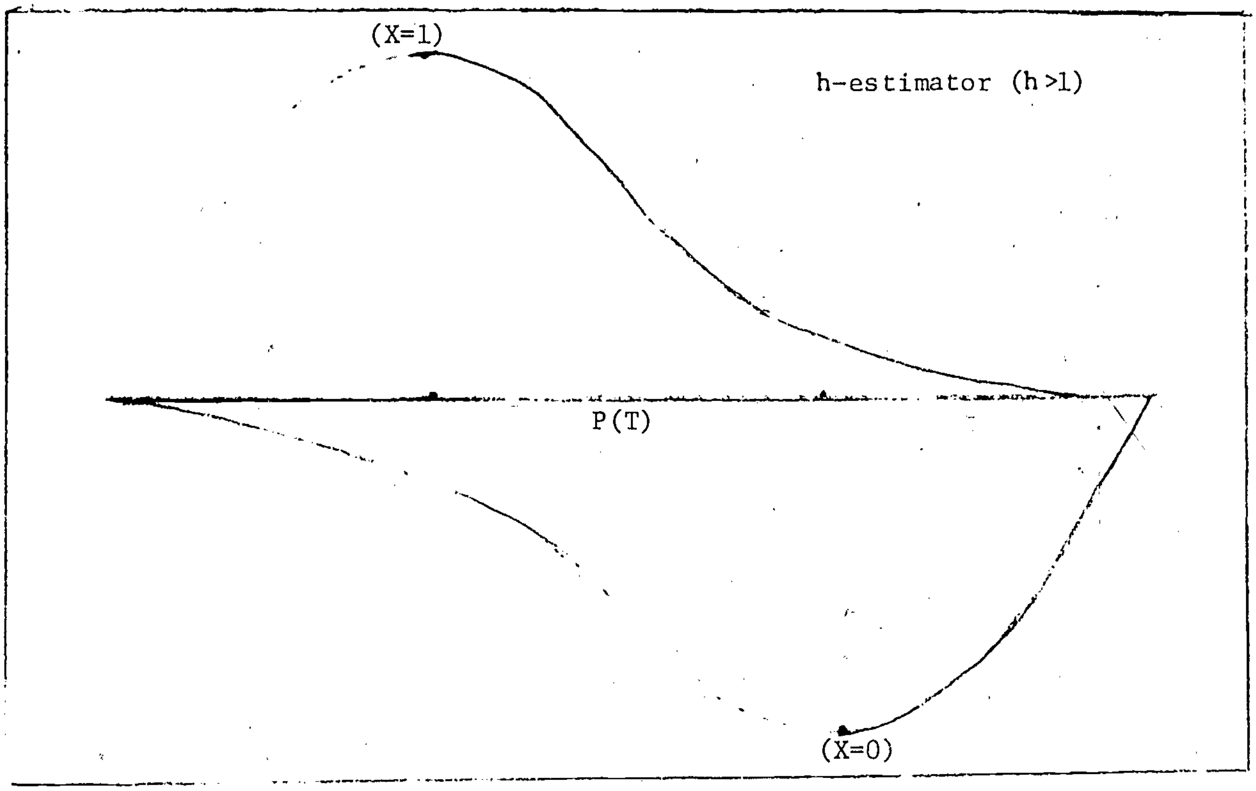
h	1.5	2.0	3.0	4.0	5.0
EFF	.99	.98	.95	.92	.90

Table 3: Efficiency comparison of h-estimators to the MLE under a neighboring model

h	1.5	2.0	3.0	4.0	5.0
EFF	2.24	4.83	19.15	51.46	82.62

8. Figures

Influence Curves



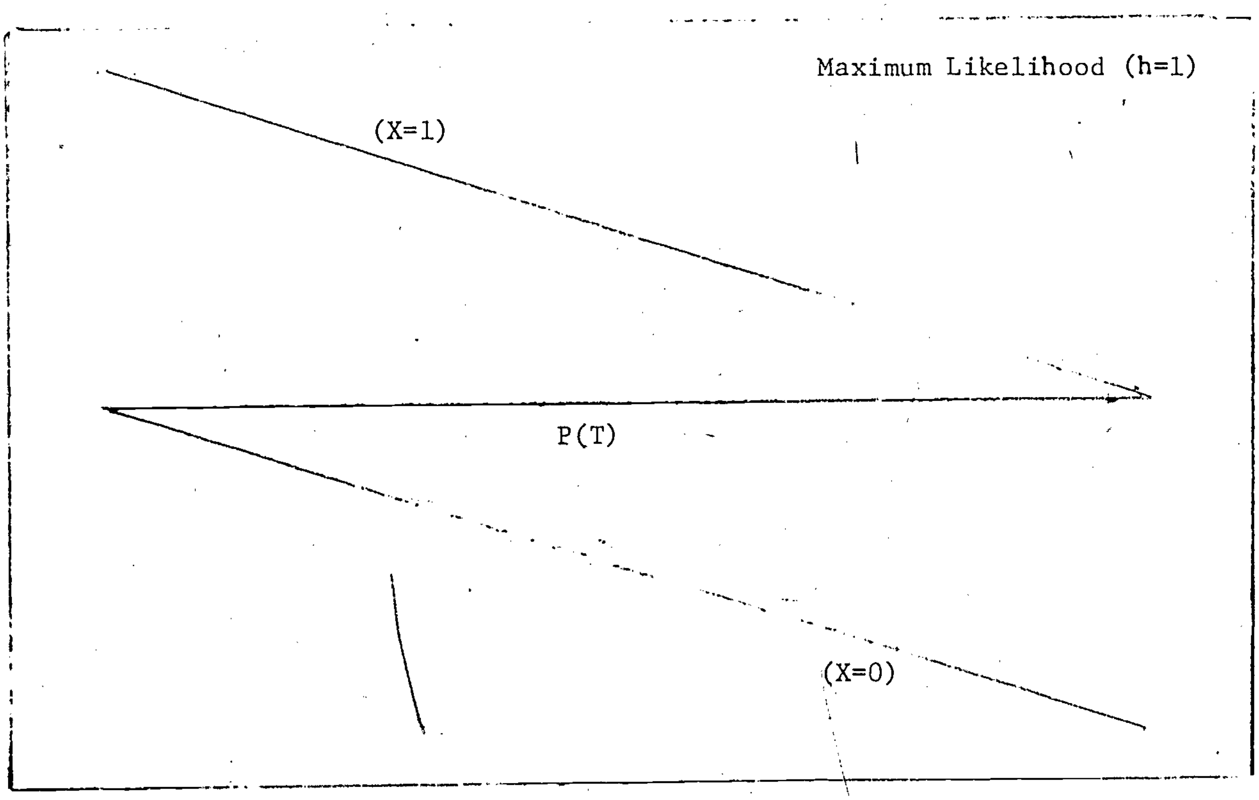


Figure 1